OCCUPATIONAL CODING: BEST CODING AND CATI CODING METHODS

Rebecca Rak, Selma Chen, and Lucinda Gray

In the 1995 National Survey of Recent College Graduates (NSRCG:95), data were collected on the respondent's primary (or current) occupation, previous occupation, and second occupation. To assure consistency in occupational coding, after a case was completed by telephone interview or mail survey, an occupational worksheet for each graduate was generated. All the occupational codes selected by the graduates were subject to review. This review was based on the work-related information provided by the graduates. If the self-selected code was inappropriate, a new or "best" code was assigned and the occupation code was updated.

The first portion of this document describes the materials used to evaluate each occupational code assignment, the reasons codes were changed, and an overview of the changes that were made for each type of occupational question (primary occupation, previous occupation, and second occupation). This is followed by a more detailed assessment of best coding assignments made for the primary occupation question. In the detailed section, the various CATI coding methods are discussed as well as the number of recodes and the timings for each method.

Occupational Worksheets

In order to facilitate the reviewing process, separate worksheets were produced for each graduate by occupational question (i.e., primary, previous, and second). For the primary job (B17), these worksheets included occupational codes and job descriptions, data related to the graduates' education, the relationship between education and work, employer information, salary data, and major work activities. For the previous (B9) and second (B37) occupations, only the respondent's major field of study and job description were included. In addition, all the comments made by the telephone interviewers were reviewed. All occupational worksheets were produced on a weekly basis.

Updating the Occupational Codes

For each occupation question, two coding variables were maintained. The first variable contained the self-selected code chosen by the graduate during the interview or on the mail survey. The second variable contained the updated or "best" code assigned by the occupation coder. These updated codes were entered into the database by the coder using a specially designed computer program that included the appropriate edit checks.

Types of Coding Changes

Upon receipt of occupational worksheets, all the self-selected occupational codes and relevant data were carefully reviewed to see if these self-selected codes were appropriate. If the self-selected code was not the most appropriate code, then an updated or "best" code was assigned and entered into the database. These coding changes can be divided into four groups. The first group includes the handling of refusals and "don't know" responses. The second group includes changes made to follow the new NSF occupation coding guidelines. The third group involves the changes from "other" categories to more specific categories, and the fourth group includes changes made to correct inappropriate self-selected codes.

Refusals and don't know. When the respondent was unable to chose an appropriate occupational code or refused to select an occupational code, but provided job descriptions and/or answered all the work related and other relevant questions, the occupational worksheets were reviewed to determine if a best code could be assigned to the occupation. If there was sufficient information to assign a best code, a valid code was assigned; otherwise, the case was sent to data retrieval.

New NSF guidelines. During data collection, NSF developed several new occupation coding guidelines. These changes were incorporated into the best coding process. Below is a list of four changes that were frequently made to adhere to these new standards.

- **Teaching in a non-school setting.** Westat developed a 501 code for teachers in nonacademic settings (e.g., continuing education teachers and computer training instructors for large corporations). In accordance with NSF guidelines, however, most of the codes in this category were recoded to 257 or to other teaching codes.
- Other craft and trade occupations. During the telephone interview, the 406 code was used to classify craft and trade occupations that did not fit into the 401-405 codes. Since this code was not included in the NSF occupational list, all of these were recoded.
- Sales managers and supervisors. In the 1993 National Survey Recent College Graduates (NSRCG:93), first line supervisors and managers in sales and marketing occupations were classified in the same category as the workers they supervised. In the 1995 survey, however, they were recategorized under the 203 code to be compatible with the new NSF guidelines.
- **Recreational workers and athletes.** In the 1993 survey, recreational workers were categorized as social workers (240), but in the 1995 survey this group was recategorized under 500 (other occupations). Athletes were also recategorized, moving from 010 to 500.

Changes from "other" categories to more specific categories. In general, occupational codes can be divided into two groups. The first group is referred to as the "non-other" occupational classifications. These classifications include specific occupations that are contained under broad occupational headings. For example, statistician is a specific occupation located under the broad heading of mathematical scientists. The second group of codes is referred to as the "other" codes. Each broad heading (e.g., mathematical scientists) contains an "other" category (e.g., other mathematical scientists) for occupations that are related to the broad heading but would not be appropriately classified under one of the specific occupations.

Graduates usually selected an "other" code when they were able to narrow their coding selection down to one broad occupation category (e.g., mathematics) but could not select a more specific code within the broad category. For example, a university research assistant in mathematics may have coded himself/herself under "other mathematical scientists," or a systems analyst may have selected the "other computer scientists" category. Many of these cases were moved from "other" classifications during best coding.

An "other" category, however, does not necessarily imply a "non-specific" category. Sometimes "other" codes include specific occupations, such as environmental biologist or pilot. Thus, instead of changing codes from an "other" category to a "non-other" category, some codes were updated to an "other" category code. For example, the 500 code encompasses a variety of occupations, which are not included under any other codes (e.g., recreation workers, athletes, soldiers, helpers, laborers, etc.). Therefore, instead of changing codes from 500 to a more specific code, some occupations were best coded to 500. Similar examples can be found for environmental scientists. Environmental scientists were best coded to 027 (other biological scientists) or 198 (other chemical scientists).

Reviewing self-selected codes. In some cases, comparing the self-selected code with the open-ended job descriptions, probes, and all the relevant information, revealed that the respondent had selected an inappropriate occupation code. These selection errors may have resulted because respondents misinterpreted or misunderstood the question, based their selections on their occupational self-perception and these perceptions differed from NSF coding standards, or selected a code that was somewhat appropriate given the occupation but was not the best choice. The first explanation for errors is difficult to identify and could be related to a multitude of factors, including respondents failing to listen to the question, interviewers failing to probe extensively, or respondents being unable to accurately answer questions about a job they may have left shortly after April 15. Different self-perceptions of occupations and the coding categories, however, often led respondents to select an inappropriate job code. For example, many restaurant and store managers selected the 141 code (top and mid-level managers) without also listing supervising or managing as a primary or secondary job activity, and/or without earning a relatively high salary. In many of these cases, respondents based their coding choices on their job titles rather than occupational responsibilities. The third explanation, selecting a somewhat appropriate code, refers to situations where a respondent selected a code that seemed appropriate to the respondent, but based on the NSF coding system actually belonged elsewhere. For example, a respondent who taught mechanical engineering at a university may have classified himself/herself in the engineering occupations but according to the NSF occupation coding system would be more accurately classified under teaching occupations.

Best Coding Results

As previously described, all occupational responses were reviewed by an occupation coder to determine whether the self-selected code was the most appropriate code. Data were collected to determine how responses were recoded. Table 1 examines the self-selected codes according to whether the respondent selected an "other" code or a "non-other" code during the telephone interview or on the mail survey, and how many of each were recoded. Table 2 reviews the updated codes that were assigned to occupations to determine how many were moved into "other" categories and how many were moved into "non-other" categories. A total of 17,894 responses were received to occupation questions. The majority of the responses were received for the primary job question (14,068 responses) with 2,117 responses to the previous job question, and 1,709 to the second job question. Of the total responses, 4,444 (24.8 percent) received updated codes during the recoding process.

Table 1 describes which self-selected responses were recoded based on the type of code (other/non-other) they originally received. Among all self-selected responses, 16.4 percent were recoded from the "other" category, and 8.5 percent were recoded from the "non-other" category. For the primary job question, the percentages of recodes mirrored the overall totals, with 24.6 percent recoded (16.0 percent from "other" and 8.6 percent from "non-other"). The percentage of recodes was slightly higher for the second job (28 percent), but this item included a much smaller number of responses.

(**************************************			8						
Recode status and self-selected	To (all j	tal obs)	Prima (B)	ry job 17)	Previo (B	ous job 19)	Second job (B37)		
code status	Number	Percent	Number	Percent	Number	Percent	Number	Percent	
Total occupation code responses	17,894	100%	14,068	100%	2,117	100%	1,709	100%	
Total responses recoded	4,444	24.8	3,458	24.6	505	23.9	481	28.1	
Self-selected code status "other" Self-selected code status	2,930	16.4	2,255	16.0	364	17.2	311	18.2	
"non-other"	1,514	8.5	1,203	8.6	141	6.7	170	9.9	
Total responses not recoded	13,450	75.2	10,610	75.4	1,612	76.1	1,228	71.9	

Table 1.Total occupation code responses by job type, recode status, and self-selected code status
(other/non-other) prior to best coding1

¹Self-selected code status prior to best coding refers to the occupation code that was selected by the respondent during the CATI interview or on the mail survey. "Other" status indicates that the self-selected occupation code was assigned in a classification labeled "other" such as "other mathematical scientist" or "other biological/life scientists." The "non-other" status indicates that a specific occupation code (e.g., statistician or biochemist) was assigned.

NOTE: Totals include all occupation responses received to CATI interviews and mail surveys. Because of rounding, details may not add to totals.

SOURCE: National Science Foundation/SRS, National Survey of Recent College Graduates, 1995

Responses were also analyzed based on the type of updated code (other/non-other) that was assigned during best coding. Of the total responses, 8.2 percent were updated to an "other" classification and 16.6 percent were updated to a "non-other" classification (table 2). Once again, the distribution of primary job updates was nearly identical to these changes (7.8 percent and 16.7 percent, respectively). For the previous job, 8.0 percent of the responses were updated to an "other" classification and 15.8 percent were updated to a "non-other" classification. The second job responses had the highest percentage of updates to an "other" classification (11.5 percent) and the same percentage of updates to a "non-other" classification (16.7 percent) as the primary job responses.

 Table 2.
 Total occupation code responses by job type, recode status, and updated code status (other/non-other) after best coding¹

	/		0						
Recode status and	То	tal	Prima	ry job	Previo	ous job	Second job (B37)		
updated	(all j	obs)	(B	17)	(B	9)			
code status	Number	Percent	Number	Percent	Number	Percent	Number	Percent	
Total occupation code responses	17,894	100%	14,068	100%	2,117	100%	1,709	100%	
Total responses recoded	4,444	24.8	3,458	24.6	505	23.9	481	28.1	
Updated code status "other"	1,468	8.2	1,102	7.8	170	8.0	196	11.5	
Updated code status "non-other"	2,976	16.6	2,356	16.7	335	15.8	285	16.7	
Total responses not recoded	13,450	75.2	10,610	75.4	1,612	76.1	1,228	71.9	

¹Updated code status after best coding refers to the occupation code that was selected by the data preparation staff as being more accurate or appropriate than the self-selected code. "Other" status indicates that the updated occupation code was assigned in a classification labeled "other" such as "other mathematical scientist" or "other biological/life scientists." The "non-other" status indicates that a specific occupation code (e.g., statistician or biochemist) was assigned.

NOTE: Totals include all occupation responses received to CATI interviews and mail surveys. Because of rounding, details may not add to totals.

SOURCE: National Science Foundation/SRS, National Survey of Recent College Graduates, 1995

CATI Occupational Coding Methods

During the telephone interview, each occupation was assigned a code in one of three ways: autocoding, new branching, or the standard method. The main difference between these methods was the number of occupation screens that were read to the respondent. Each coding method is discussed in greater detail below.

Autocoding. Autocoding was a process by which the respondent's occupation was coded on the basis of a verbatim or probe response. CATI read each response line-by-line, beginning with the second probe and working toward the initial verbatim response while looking for an exact match to an occupation contained in a specially designed occupation dictionary. The occupation dictionary was compiled based on occupation responses given on the 1993 cycle of NSRCG. The 1993 responses were examined to identify those that received distinct occupation codes. For example, accountant, tax accountant, and certified public accountant generally received a 151 code. These occupations were placed in the autocode portion of the dictionary with the corresponding NSF code. When CATI found an exact match to one of these occupational entries, the appropriate code was automatically assigned to the respondent's occupation. In these cases, the respondent did not go through any additional occupation coding screens, and, consequently, did not actually select a code. For analysis purposes, however, the code assigned through the autocoding process is still considered self-selected, since the code assignment was made during the interview based on information provided by the respondent.

New branching. New branching was a system that bypassed the first standard occupational category screen (i.e., a screen that lists 9 broad categories from which the respondent picks one to begin a potentially lengthy branching sequence resulting in a code selection). Eight "new" branches were designed to reduce occupation coding time. These branches were based on occupations that could not be assigned a unique code but were identified as belonging in one of two or three broad categories. For example, a respondent working in the computer industry would be asked to classify his/her occupation as either a computer occupation, an engineering occupation, a clerical occupation, or something else. If one of the first three categories was selected, CATI would go to the appropriate broad heading screen. If the "something else" option was selected, CATI would return to the standard occupational category screen, and proceed with the standard coding method described later in this section. Three branches led to an abbreviated listing of broad categories, such as in the aforementioned case. Thus, the interviewer only read three or four options to the respondent, and an entire branching step was eliminated (i.e., this is a comparison to making the same choice under the standard method). For the remaining 5 branches, CATI proceeded directly to a subcategory screen (i.e., a screen listing options under a broad heading); thus, bypassing one or more branches in the process. Each of the eight branching schemes was assigned a letter code (as opposed to an NSF occupation code) in the occupation dictionary.

Standard coding method. The third coding option used was the standard method. This coding scheme was developed during the 1993 survey. The interviewer read a list of nine broad categories to the respondent. The respondent selected the most appropriate category, and continued down a specific branching path.

Self-selected response codes. Of the 13,357 final responses received to the primary occupation question during the telephone interview, 22.0 percent of the responses were autocoded, 18.5 percent went through the new branching scheme, and 59.5 percent were coded using the standard method (see table 5).

Best Coding Based on CATI Coding Method

This section contains an analysis by coding method of primary job question responses that required recoding during the best coding process. Table 3 describes which self-selected codes were recoded based on the original classification type (other/non-other) and whether the response was autocoded, went through a new branch, or went through the standard method. Only 2.1 percent of all autocoded responses were recoded, with 1.1 percent recoded from "other" classifications and 1.0 percent recoded from "non-other" classifications. For new branching, only 15.4 percent of the self-selected responses were recoded. These recodes were almost evenly divided between "other" and "non-other" classifications (8.1 percent and 7.4 percent, respectively). Among responses that were coded using the standard coding method, 35.9 percent were recoded during the best coding process. This was a substantially higher percentage of responses than for either autocoding or new branching. For the standard branching method did not simplify the coding process for the respondent by automatically assigning an occupation code or eliminating extraneous broad categories. In addition, occupations that went through the standard branching scheme were generally those that were not easily classified; otherwise they would have been autocoded or included in the new branching.

Table 3.	Primary job responses (B17) by CATI coding method, recode status, and self-selected	ima	Prima	Prima	imary	y jo	ob :	res]	spons	es	(B17) b	y	CATI	codi	ng	method	, re	code	statu	ıs,	and	S	elf-se	elect	ed
	code status (other/non-other) prior to best coding ¹	de s	code s'	ode st	le sta	atus	s (o	othe	ier/no	n-o	other) p i	ria	or to b	est co	odiı	ng ¹									

		/ L							
Recode status and self-selected	To (B)	tal 17)	Autoc	coding	New br	anching	Standard		
code status	Number	Percent	Number	Percent	Number	Percent	Number	Percent	
Total occupation code responses	13,798	100%	3,051	100%	2,526	100%	8,221	100%	
Total responses recoded	3,408	24.7	63	2.1	390	15.4	2,955	35.9	
Self-selected code status "other" Self-selected code status	2,212	16.0	33	1.1	204	8.1	1,975	24.0	
"non-other"	1,196	8.7	30	1.0	186	7.4	980	11.9	
Total responses not recoded	10.390	75.3	2.988	97.9	2,136	84.6	5.266	64.1	

¹Self-selected code status prior to best coding refers to the occupation code that was selected by the respondent during the CATI interview. "Other" status indicates that the self-selected occupation code was assigned in a classification labeled "other" such as "other mathematical scientist" or "other biological/life scientists." The "non-other" status indicates that a specific occupation code (e.g., statistician or biochemist) was assigned.

NOTE: Totals only include CATI responses to the primary job question. Because of rounding, details may not add to totals.

SOURCE: National Science Foundation/SRS, National Survey of Recent College Graduates, 1995

Table 4 reviews the updated codes assigned during the best coding process based on the type of classification (other/non-other) assigned and the CATI coding method used during the self-selection process. For autocoded responses, only 1.0 percent were recoded into an "other" classification and 1.1 percent received a "non-other" classification. New branching recodes were generally classified in the "non-other" categories, but for the standard coding method responses, more responses were assigned to "other" codes than to "non-other" codes (24.3 percent compared with 11.7 percent). Reasons for assigning "other" codes were discussed earlier in this report.

Recode status and updated	To (B	tal 17)	Autoc	oding	New br	anching	Standard		
code status	Number	Percent	Number	Percent	Number	Percent	Number	Percent	
Total occupation code responses	13,798	100%	3,051	100%	2,526	100%	8,221	100%	
Total responses recoded	3,408	24.7	63	2.1	390	15.4	2,955	35.9	
Updated code status "other"	2,130	15.4	29	1.0	105	4.2	1,996	24.3	
Updated code status "non-other"	1,278	9.3	34	1.1	285	11.3	959	11.7	
Total responses not recoded	10,390	75.3	2,988	97.9	2,136	84.6	5,266	64.1	

Table 4.Primary job responses (B17) by CATI coding method, recode status, and updated code
status (other/non-other) after best coding1

¹Updated code status after best coding refers to the occupation code that was selected by the data preparation staff as being more accurate or appropriate than the self-selected code. "Other" status indicates that the updated occupation code was assigned in a classification labeled "other" such as "other mathematical scientist" or "other biological/life scientists." The "non-other" status indicates that a specific occupation code (e.g., statistician or biochemist) was assigned.

NOTE: Totals only include CATI responses to the primary job question. Because of rounding, details may not add to totals.

SOURCE: National Science Foundation/SRS, National Survey of Recent College Graduates, 1995

Timings

A second CATI coding issue was the speed with which responses were coded based on the coding method. The response timings were measured in two ways. First, CATI calculated the time from when the primary job question was asked through the time when an occupation code was selected. CATI also calculated the time it took to select a code after the verbatim and probe responses were collected. Overall, the mean processing time for the first process was 124.8 seconds, or slightly more than 2 minutes (table 5). The mean processing time for selecting the code only was 70.3 seconds, or slightly more than 1 minute. Autocoding posted the lowest times for both types of timings, with a mean processing time for the full process of 45.4 seconds, and a code selection time of 1.3 seconds. For the new branching method, the mean processing time for asking the occupation question and selecting the code was 108.4 seconds. The time to select the code only was approximately half that amount (56.6 seconds). The mean processing time for the full process using the standard coding method took over three times longer than that for the autocoding method (159.3 seconds compared with 45.4 seconds). The time to select an occupation code was almost 100 times greater than that for autocoding and nearly double that of new branching taking 100.1 seconds.

 Table 5.
 Mean processing time for primary job responses (B17) by timing method and CATI coding method

CATI coding method	Total responses	Full question	Occupation coding only		
		(Processing tin	ie in seconds)		
Total B17 occupation responses	13,357	124.8	70.3		
Autocoding	2,941	45.4	1.3		
New branching	2,474	108.4	56.6		
Standard	7,942	159.3	100.1		

NOTE: Totals only include CATI responses to the primary job question for final completes. Because of rounding, details may not add to totals.

SOURCE: National Science Foundation/SRS, National Survey of Recent College Graduates, 1995